



## ORIGINAL RESEARCH PAPER

## Benchmarking Machine Translation Efficacy for Teaching EAP Reading Comprehension Skills

V. Mirzaeian<sup>\*1</sup>, M. Maghsoudi<sup>2</sup><sup>1</sup> Department of English, Faculty of Literature, Alzahra University, Tehran, Iran<sup>2</sup> Department of English Language Teaching, Farhangian University, Tehran, Iran

## ABSTRACT

Received: 31 May 2023  
 Reviewed: 8 September 2023  
 Revised: 19 October 2023  
 Accepted: 04 December 2023

## KEYWORDS:

Comprehensibility  
 EAP  
 Machine Translation  
 Reading Comprehension

Corresponding author

[mirzaeian@alzahra.ac.ir](mailto:mirzaeian@alzahra.ac.ir)

(+98912) 6047452

**Background and Objectives:** Although Machine Translation (MT) is extensively researched within the field of Artificial Intelligence (AI) and translation studies, few studies have attempted to implement MT output in foreign language teaching (FLT). One potential aspect of using MT in FLT refers to the implementation of MT output for reading comprehension. Considering the existing gap in the body of research on this issue, the present study aimed to investigate whether MT output is qualified enough to be used as an aid in EAP reading comprehension courses. More specifically, this study intended to benchmark the efficacy of MT output for EAP reading comprehension courses based on the data obtained from testing its comprehensibility and probing the students' perceptions. To achieve the objectives of the study, MT was operationally defined as quality assessment in terms of output efficacy, a combination of usability and comprehensibility, which mirrors the ultimate goal of MT use in EAP reading comprehension courses, from the users' or target readers' standpoint. Within this perspective, the current research was an attempt to assess the quality of MT output in terms of comprehensibility and the degree to which MT output might be comprehensible to the EAP students participating in this study.

**Materials and Methods:** The participants of the study, 140 Iranian undergraduate university students majoring in the field of education at Farhangian University, Iran, were selected based on simple random sampling. Oxford Quick Placement Test was used to homogenize them in terms of English proficiency. Two versions of a reliable reading comprehension test, human translation (HT) and Machine Translation (MT), were given to. This test included 25 multiple-choice items, assessing the participants' literal comprehension of information stated in the passage as well as higher-order comprehension that required making inferences and conclusions. In particular, the items measured textual coherence, inference, reference, scanning, skimming, and word-meaning inference. To test the reliability of the tests, the KR-21 formula was applied and the results showed that both HT test (.83) and MT test (.78) were reliable. To investigate the perceptions of the participants on the efficacy of the MT output they encountered on the test, semi-structured interviews were conducted with some of the participants in Persian.

**Findings:** With reference to the results of non-parametric tests such as Spearman's rho, and Mann-Whitney Tests, and considering the observed effect sizes (Cohen's d), it was revealed that, generally, the efficacy of MT output is comparable to that of HT. Moreover, in terms of reading comprehension sub-skills, the qualities of the two translations were comparable with regard to scanning, and inference, but not skimming and reference. Furthermore, the findings from the interview indicated that the students perceive MT to be a seminal aid for their EAP reading comprehension activities despite the minor problems that exist in the output such as morpho-syntactic errors or inappropriate lexical equivalents.

**Conclusions:** The present study confirmed the fact that the efficacy of MT output is target-reader-dependent and text-dependent since it is determined both by the characteristics of the readers, such as their disciplines, and text features, as demonstrated by the significant differences in comprehension levels of the same readers measured by the same questions for HT and MT output. Accordingly, this study shed limelight on comprehensibility as a criterion of MT output efficacy; that is to say, it has to be reminded that MT quality needs to be defined as a context-bound and target-reader-specific concept.



## COPYRIGHTS

© 2024 The Author(s). This is an open-access article distributed under the terms and conditions of the Creative Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/>)



NUMBER OF REFERENCES

62



NUMBER OF FIGURES

0



NUMBER OF TABLES

3

## مقایسه کارایی ترجمه ماشینی برای آموزش مهارت های درک مطلب در انگلیسی برای اهداف آکادمیک

وحیدرضا میرزائیان<sup>۱\*</sup>، مجتبی مقصودی<sup>۲</sup><sup>۱</sup> گروه زبان انگلیسی، دانشکده ادبیات، دانشگاه الزهراء، تهران، ایران<sup>۲</sup> گروه زبان انگلیسی، دانشگاه فرهنگیان، تهران، ایران

## چکیده

**پیشینه و اهداف:** ترجمه ماشینی یکی از فنون هوش مصنوعی است که با استفاده از رایانه، متون را از زبانی به زبان دیگر ترجمه می کند. مطالعات محدودی با هدف استفاده از خروجی ترجمه ماشینی برای اهداف زبان آموزی انجام شده است. یکی از جنبه های استفاده از ترجمه ماشینی در آموزش زبان به استفاده از خروجی ترجمه ماشینی برای درک مطلب در یادگیری زبان اشاره دارد. با توجه به این واقعیت که کیفیت خروجی ترجمه ماشینی موضوعی بحث برانگیز بوده است، پژوهش های بیشتری برای بررسی این موضوع انجام داده شد. این پژوهش با هدف ارزیابی کارایی خروجی ترجمه ماشینی برای درک مطلب انگلیسی برای اهداف آکادمیک است، مورد نیاز است. این پژوهش با هدف ارزیابی کارایی خروجی ترجمه ماشینی برای درک مطلب انگلیسی برای اهداف آکادمیک از طریق اثبات قابل درک بودن آن و بررسی ادراک دانشجویان انجام شد. این پژوهش تلاشی برای ارزیابی کیفیت خروجی ترجمه ماشینی از نظر قابل درک بودن خروجی ترجمه ماشینی برای دانشجویان شرکت کننده بود.

**روش:** برای همگن سازی شرکت کنندگان از نظر تسلط به زبان انگلیسی، از آزمون تعیین سطح زبانی آکسفورد استفاده شد. دو نسخه از آزمون درک مطلب یعنی ترجمه انسانی و ترجمه ماشینی به ۱۴۰ دانشجوی ایرانی در مقطع کارشناسی دانشگاه فرهنگیان داده شد. این آزمون شامل ۲۵ آیتم چند گزینه ای بود که درک شرکت کنندگان از اطلاعات مندرج در متن و همچنین درک مرتبه بالاتری که مستلزم استنباط و نتیجه گیری بود را ارزیابی می کرد. به منظور بررسی پایایی آزمون ها از فرمول KR-21 استفاده شد و نتایج نشان داد که هر دو آزمون ترجمه ماشینی (۰/۸۳) و آزمون ترجمه انسانی (۰/۷۸) پایایی دارند. به منظور بررسی ادراک شرکت کنندگان از کارایی خروجی ترجمه ماشینی که در آزمون با آن مواجه شدند، مصاحبه های نیمه ساختاریافته با برخی از شرکت کنندگان انجام شد.

**یافته ها:** از طریق آزمون های ناپارامتریک مانند آزمون های اسپیرمن و من ویتنی مشخص شد به طور کلی، خروجی ترجمه ماشینی با خروجی ترجمه انسانی قابل رقابت است. علاوه بر این، از نظر مهارت های فرعی درک مطلب، کیفیت دو ترجمه قابل مقایسه بود. علاوه بر این، یافته های حاصل از مصاحبه نشان داد دانش آموزان علی رغم مشکلات جزئی موجود در خروجی مانند خطاهای صرفی - نحوی یا معادل های واژگانی نامناسب، ترجمه ماشینی را کمکی اساسی برای فعالیت های درک مطلب خود می دانستند.

**نتیجه گیری:** این پژوهش نشان داد کارایی خروجی ترجمه ماشینی وابسته به خواننده هدف و وابسته به متن است، زیرا هم توسط ویژگی های خواننده و هم متن تعیین می گردد. این موضوع در نمرات کسب شده توسط دانشجویان در گروه های مختلف کاملاً مشهود بود. بنا براین پژوهش حاضر درک پژوهشگران را از ترجمه ماشینی و تاثیر آن بر درک مطلب افزایش می دهد.

تاریخ دریافت: ۱۰ خرداد ۱۴۰۲

تاریخ داوری: ۱۷ شهریور ۱۴۰۲

تاریخ اصلاح: ۲۷ مهر ۱۴۰۲

تاریخ پذیرش: ۱۳ آذر ۱۴۰۲

## واژگان کلیدی:

ادراک پذیری

انگلیسی آکادمیک

ترجمه ماشینی

درک مطلب

\* نویسنده مسئول

mirzaeian@alzahra.ac.ir

۰۹۱۲-۶۰۴۷۴۵۲

## Introduction

The contribution of MT to language learning has not been limited to the realm of general English instruction but to teaching English for academic purposes (EAP). This phenomenon is partly rooted in the fact that translation and L1 use have been dominantly used in EAP classes, especially those held in EFL contexts since long ago [1,2]. Considering the growing use of MT in

EAP courses and by higher education students, there has been a growing body of research evaluating MT output use in EAP courses or in higher education settings, in general. For example, Groves and Mundt [3] focused on a routine academic activity of essay translation. Analyzing the grammatical errors in the MT output, they found that MT was unable to produce an error-free text. However, in comparison to international testing standards,

the accuracy of the output was almost the minimum required for university admission. Bahri and Mahadi [4] also reported that many higher education students consider Google Translate (GT) as an optimal supplementary learning tool for improving their vocabulary, reading, and writing which encourages independent study self-efficacy. Lee [5] also warned against the limitation of MT use for language teaching in academic settings and cautioned EAP teachers to provide sufficient guidance to their students while benefiting from the constructive and positive aspects of MT in their instructional programs.

However, it is worth mentioning that the results of the previous studies were somewhat controversial in terms of the merits of MT output use for English by higher education students. On one hand, some studies such as Tsai [6], advocated the use of MT output in EAP courses due to its higher writing quality in terms of fewer spelling and grammar errors, and because of students' gratification with GT in their English writing course, especially in terms of choosing appropriate word and writing task completion. On the other hand, some researchers criticized MT use in higher education English courses. Suhono et al. [7], for instance, showed that GT provided low-quality and ineffective English-Indonesian sentences, assessed in terms of lexico-grammatical and textual equivalence. Knowles [8] also alerted about the need for rethinking higher education English learners' use of MT, EAP teachers' MT integration and helping them access MT training.

According to what has been discussed so far, MT is like a double-edged sword whose usefulness is not easy to determine. Moreover, most of the previous studies conducted on the use of MT in EAP courses concentrated on developing the learners' writing skills. Accordingly, there seems to be a gap with

regard to the use of MT as an educational aid in teaching reading comprehension skills in EAP courses.

## Review of the Related Literature

---

### The Role of Translation in EAP Instruction

Using L1 has been a debatable issue in EFL instruction and English for specific purposes (ESP) teaching and it can be traced back to four decades ago [1]. Accordingly, it has been a cannon of several studies. In the same vein, as Rushwan [9] stated, translation has been considered a pedagogical tool in EAP courses for teaching reading comprehension. It seems that the main reason for relying on translation lies in the fact that most higher education students in EFL contexts face bewilderment and difficulties in comprehending technical texts in English due to their low level of English proficiency and limited vocabulary knowledge [9,10]. Thus, they resort to translation to handle the large body of texts they need to absorb.

Higher education students' tendency to use L1 and translation as a shortcut to technical text comprehension, has led some ESP experts to support the notion that language learning can be greatly facilitated via the implementation of the mother tongue while teaching ESP. [11]. They have also provided evidence on the effective role of using L1 and translation teaching EAP materials [11,12]. For example, the studies by Malekan and Hajimohammadi [13] and Pham [14] revealed a significant relationship between EAP learners' translation ability and their resilience in reading comprehension as well as their reading comprehension scores. Some researchers such as Tavakoli et al. even claimed that "translation has a high potentiality to work as a reliable and valid tool to assess reading comprehension" since "there exists a high positive correlation between the participants' proficiency in reading

comprehension and their proficiency in translation" [15, p. 93].

Furthermore, translation is seen as a proper teaching technique in EAP courses [16]. To advocate this standpoint some researchers argued that translation tasks can provide EAP teachers with a teaching methodology to guide their learners in comprehending and producing texts that feature the quality criteria of stylistic fluency and terminological accuracy. In the same line, Leonardi [17], Mažeikiene [18], and Novita and Mustafha [19] also indicated that translation tasks consolidate targeted language skills of EAP learners, facilitate interaction and comprehension of technical texts, and develop learners' analytic skills and language accuracy.

Using electronic tools such as electronic dictionaries is common among EFL learners and many of them proved to be aware of the potential and limitations of such educational aids [20]. MT is also one of these electronic devices that has long been researched as an educational aid, despite its limitations and output deficiencies. Undeniable advances in the accuracy of MT outputs have considerably increased its use as a pedagogical tool to support EAP instruction [21]. It seems that higher education students of different English language proficiency enjoy its benefits, especially, when it comes to writing in English academically [21, 22], and MT is not limited to the field of translation but serves as a translation aid in academic and pedagogical contexts [23].

The evidence from the related research on using MT in language teaching testifies to the fact that the learners of both higher and lower proficiency levels in L2 can benefit from a variety of MT engines, even free ones such as GT to cater to their needs despite their limited command of academic and technical terms [24]. MT mediation, in spite of its imperfect output, encourages more engagement with language

learning tasks which leads to better learning outcomes [5,24]. In addition to the observed performance of the L2 learners after MT-based instruction, the investigations of their perceptions also confirm the satisfactory role of MT in learning EAP, especially in terms of looking up lexical items and improving English writing [6]. Interestingly, less proficient EAP learners proved to be more satisfied with the contribution of MT to their learning and achievements in using English for academic purposes [6].

Considering the expanded access to online translation services and globalization which raises the demand for such translation services, most of the existing research on the use of MT in language teaching has coped with MT post-editing or MT output correction which is intertwined with multilingual communication of professionals [22]. They have indicated that online MT is of great educational value in L2 writing instruction and, generally, report L2 learners' improvements after MT-based instruction in terms of syntactic complexity, accuracy, lexical complexity, and fluency [25]. Despite the scarcity of research on the contribution of MT to the reading comprehension ability of EAP learners, there is some promising evidence of the helpfulness of MT in EAP instruction as the learners find it useful, easy to use, and satisfactory [26].

### **MT Output Quality Assessment and MT Efficacy**

It seems that the emergence of MT has added to the complexity of defining and operationalizing translation quality assessment. The evaluation of MT systems is a complex task. Several approaches, orientations, and perspectives have been adopted to evaluate MT outputs which have resulted in various operational definitions of MT translation quality assessment. A bird eye view of the previous

studies [27,28] disclose two major trends, human evaluation and non-human evaluation, both of which were more linguistically than cognitively oriented. In addition to their methodological aspects, as Alhaisoni [28] affirmed, MT evaluation pursues three main purposes: (1) error analysis which aims at detecting and analyzing possible cases of errors; (2) system comparison which entails measuring the effectiveness of the MT system and occasionally involves comparison of various versions of a system or ultimate versions of different systems. For the sake of this purpose, quality assessment may require comparison of translations by different sources or systems; (3) system optimization which is mainly conducted for adjusting internal parameters of MT systems to maximize MT system quality.

Human evaluation has been commonly criticized for being subjective, therefore, the reliability of their translated products has been questioned accordingly [29]. Moreover, it has been blamed for being expensive and time-consuming [30]. In addition, with regard to the fact that there are thousand-page and million-word corpora to be assessed, human evaluation seems impractical for most purposes [31]. To reduce problems and provide a practical procedure, automatic MT evaluation is introduced which mainly uses similarity metric to assess sentence closeness between the MT output and its set of references [32, 33, 34]. However, there is still a concern about automatic MT evaluation in terms of the correlation of its results with those of human evaluation [35, 36].

The issue of MT quality assessment through human evaluation which has been criticized for its subjectivity, according to Siregar [29], is being made more objective by adopting one of the following ways. The first solution is implementing Likert scales which makes it feasible to score an output, or a document to be

more specific, at sentence and word levels. However, some experts such as Siregar [29] stated that document-level MT quality assessment is awry if the output is evaluated in this way. Simply put, machine-translated texts may have errors that go beyond individual words or sentences. For instance, there could be issues with the overall coherence of the text. As a solution, some researchers suggest using a reading comprehension test or corpus to evaluate MT quality as a whole, rather than just focusing on individual segments. This would provide a more comprehensive and objective assessment of the effectiveness of a machine translation system. In simple words, the premise is that as long as target readers of an MT output can answer the reading comprehension questions developed based on the translated texts correctly, the MT system is translating well and the output enjoys a desirable quality [37].

The current shift to comprehensibility of MT output is the byproduct of a research paradigm that focuses on the target users; however, it is undeniable that this line of research has not been much focused on. Among the few studies that have been conducted so far, some defined MT output evaluation to be equal to measuring the comprehensibility and usefulness of the output. These studies indicated that comprehensibility measured by reliable reading comprehension is certainly a valid MT output evaluation method [29, 31, 38]. Accordingly, in terms of MT output evaluation, quality assessment may be simply defined as assessing the level of the informativeness of the output for the target readers which is simply measured by the number of right answers for the comprehension questions.

Whilst a number of researchers such as Jimenez [39], focused on target readers' levels of literacy and others such as Abdelaal and Alazzawie [40], compared MT systems, the

current study is aimed at the issue of comprehensibility of MT output and investigated document-level quality according to the aspects of comprehensibility of the texts and the target readers level of comprehension. This is in line with the objective cognitive human evaluation approach of evaluating MT quality by analyzing the cognitive processes of human evaluators as they read and comprehend an MT output. This approach seeks to measure how well the machine-translated text conveys the intended meaning and how easy it is to understand for native speakers of the target language, rather than just focusing on specific grammatical or lexical errors. Objective cognitive human evaluation involves using various metrics, including response time, reading comprehension scores, and eye-tracking data, to assess the quality of the MT output. This approach is considered more reliable and effective than relying solely on subjective evaluations of human experts or users.

Although in terms of defining MT quality assessment operationally as measuring the comprehensibility of the output for target readers, the current study perpetuates the line of studies conducted by Abdelaal and Alazzawie [40], Roturier [41] Siregar, [29] Toral and Sanchez-Cartagena [38], among others, this study extends the scope of previous studies to assess the output quality in terms of the level of comprehensibility in terms the reader's purpose or need. Accordingly, the concept of translation efficacy is introduced to refer to the fact that the level of comprehension a reader needs to achieve also has to be included in judging the quality of an MT output.

MT has a significant impact on language learning, and its cognitive aspects are essential for understanding its effectiveness. It improves language learning by providing learners with instant translation of words, phrases, and

sentences in real-time, which makes it a beneficial tool for language learners [60]. Furthermore, it offers multiple benefits such as improving reading comprehension and vocabulary building, which contribute to effective language learning [61]. MT can also help in reducing cognitive load, making it easier for learners to focus on language acquisition tasks [62]. In contrast, the overuse of MT can have negative impacts on language learning, such as reducing the ability to learn grammar and syntax and weakening one's ability to engage in spontaneous communication. Therefore, learners must use MT as a tool but not as a substitute for human interaction and learning.

According to what has been discussed so far, the current study may be considered a breakthrough in terms of evaluating MT output using an objective cognitive human evaluation approach, emphasizing the process going in target readers' minds rather than the product offered by an MT system. As it is being discussed above, the present study defined MT quality assessment in terms of output efficacy which is a combination of usability and comprehensibility, which mirrors the ultimate goal of MT use in EAP reading comprehension courses, from the users' or target readers' standpoint. This research was an attempt to assess the quality of MT output in terms of comprehensibility and the degree to which MT output might be comprehensible to the EAP students participating in this study. This approach was undertaken to benchmark the implementation of MT rather than human translation (HT) in EAP courses. Having this objective in mind, the following research questions were formulated:

- Is the efficacy of MT output for EAP reading comprehension comparable to that of error-free human translation (HT) output in terms of comprehensibility?

- How do end-MT users perceive the efficacy of the MT output for EAP reading comprehension?

## Method

### Participants

The participants of this study included 140 Iranian undergraduate university students of education at Farhangian University. They were randomly selected from male and female students attending the EAP course for the students of education based on simple random sampling, i.e. a list of all the students was created, and participants were selected randomly using a random number generator. Their ages ranged from 20 to 23. The researchers outlined the research objectives and the intended use of the collected data to the participants, along with their expectations, enabling the students to make an informed decision about whether or not to participate in the study. They were informed about the researchers' expectation to take part in two exam sessions and to spend the required time committedly. In terms of their English proficiency, they were homogenized according to their Oxford Placement Test (OPT) scores; that is, the students whose scores were one standard deviation below and above the mean score were asked to participate in this study. That is the participants were selected as the homogeneous sample of this study out of the 187 students who had already taken the test. They were randomly divided into two relatively equal groups of 70 students. One group took the HT reading comprehension test and the other one took the MT one.

### Instruments

#### *English Proficiency Test*

Oxford Placement Test was used to homogenize the participants in terms of English proficiency. It is a standardized 60-item multiple-choice test

including grammar, vocabulary, and reading subsections. The participants were allotted 60 minutes to take the Oxford Placement Test. The test was found to be reliable in this study according to the calculated Cronbach's alpha (0.88).

#### *Reading Comprehension Tests*

Participants, in each group were requested to answer the questions of a reading comprehension test with two passages of similar length, topic, and text difficulty. The function of this test in the current study was to probe the effect of the type of translation output, i.e. HT or MT, on the reading comprehension of two groups of participants with similar language proficiency. Accordingly, two parallel formats of the test were developed. The first format included the English text along with its equivalent human-translated text in Persian. The second format included the English text along with its equivalent machine-translated text in Persian. The MT output was drawn from GT and inserted into the test without any modification. The rationale behind selecting this MT system was its popularity among Iranian students.

The rationale behind the availability of the English text was to create a testing environment that matches the target language use (TLU) domain of using either HT or MT output by higher education students. As argued by Alderson [42], reading assessment aims at knowing how well readers read in the real world. Accordingly, authenticity is an important feature of testing reading comprehension since it defines the link between the test and the real world. To meet these criteria, the researchers had to use texts that had not been simplified and tasks simulating real-world tasks [43]. Considering Bachman and Palmer's [44] definition of authenticity, as the degree of correspondence of a given language test task to

the features of a TLU task, and their definition of TLU domain as the situation or context in which the test taker will be using the language outside of the test itself, the researchers simulated the use of HT and MT output by Iranian higher education where they access both original and translated texts if they are required.

The entire reading comprehension test lasted 40 minutes. This test included 25 multiple-choice items, assessing the participants' literal comprehension of information stated in the passage as well as higher-order comprehension that required making inferences and conclusions. In particular, the items measured textual coherence (2 items), inference (7 items), reference (1 item), scanning (7 items), skimming (3 items), and word-meaning inference (5 items). To test the reliability of the tests, the KR-21 formula was applied and the results showed that both the HT test (.83) and the MT test (.78) were reliable.

### *Semi-structured Interviews*

To investigate the perceptions of the participants of the efficacy of the MT output they encountered on the test, semi-structured interviews were conducted with nine participants. The interview questions were generated based on the research objectives and the related literature on MT and language learning. As the study aimed to investigate the participants' perceptions of the efficacy of the MT output, the questions were designed to explore the advantages and limitations of MT in language learning. Additionally, the questions aimed to assess the accuracy and reliability of MT output and its impact on language learning progress. The questions were also created to elicit detailed and informative responses from the participants through open-ended inquiries. Finally, the aim was to generate a natural

conversation that would provide insights into the participants' experiences with MT in language learning.

It has to be noted that semi-structured interviews are a suitable qualitative research method for exploring the effect of MT on language learning. Here are some reasons why: 1) They offer flexibility, allowing the researchers to explore topics in-depth while also leaving room for unexpected insights or ideas. The questions in a semi-structured interview are usually open-ended, which allows participants to share their thoughts in their own words. 2) They tend to generate rich data due to their open-ended nature. In these interviews, participants are encouraged to provide detailed descriptions of their experiences with MT and how it has affected their language learning. This can provide valuable insights and ideas for further research. 3) They can help us gain a personal perspective on how MT is used in language learning and its effect on the participants. These interviews allow us to get a sense of how individual learners perceive and experience MT, which can help inform the development of new language learning tools or strategies. 4) They allow the researchers to explore interesting or uncertain issues more deeply by asking follow-up questions to clarify or elaborate responses from the participants. This can lead to the discovery of valuable information that may have been missed in a more structured interview.

The interviews were held with nine volunteering students on a one-to-one basis, right after the test. The interviews started with the researchers' prompts which were inspired by the previous studies and covered the students' level of satisfaction with the MT output in terms of its comprehensibility, contribution to the comprehension of the English texts, its ambiguities, and above all, its extent of usability for comprehending the texts.



In this study, thematic analysis was used which involved several steps:

- Familiarization: Researchers immersed themselves in the data by listening to interviews multiple times to gain a deep understanding of the content;
- Coding: Researchers systematically assigned codes to segments of the data that represented important concepts, ideas, or patterns;
- Generating initial themes: Codes were reviewed and grouped to identify initial themes. These themes captured the essence of the data and reflected the common patterns or meanings;
- Reviewing and defining themes: The identified initial themes were refined, defined, and named based on their relevance and coherence with the data;
- Searching for alternative explanations: Researchers critically analyzed the themes, considering alternative interpretations or explanations for the data;
- Reviewing themes against the data: The themes were reviewed in relation to the entire dataset to ensure they accurately represent the data as a whole;
- Defining and naming themes: The final themes were clearly defined, described, and labeled to provide a comprehensive understanding of the data.

The interviews were held in Persian, the students' mother tongue, and lasted for almost 15 minutes. Table 1 shows the demographic information of the participants (age, educational degree, gender using pseudonyms). The process of validating interview questions was similar to validating any research instrument and involved several steps. Before developing interview questions, it was important to have a clear definition of the research question or topic being investigated. This would help ensure that the questions are relevant and valid to the research objective.

Based on research questions, a preliminary set of interview questions was developed that was relevant and comprehensive. This was done by reviewing previous literature, consulting experts in the field, or conducting pilot interviews. A small sample of participants were asked to answer the interview questions to assess the relevance, clarity, and comprehensiveness of the questions. Based on feedback from the pilot interviews, the questions were revised and refined.

**Table 1: Participants' demographic information**

Pseudonym	Age	Gender
Reza	18	Male
Azam	21	Female
Rahim	19	Male
Diana	20	Female
Davood	22	Male
Fatima	21	Female
Hosein	18	Male
Nazila	20	Female
Jafar	21	Male

### Procedure

The study began with preparing the data collection instruments described above. The placement test was a ready-made valid instrument. However, the challenge was to develop two parallel tests of reading comprehension introduced in the previous section. It is worth mentioning that before the administration of the reading comprehension tests, they were piloted for clarity, simplicity, and time allotment. In addition, the validity of the tests was confirmed through expert judgment since four TEFL professors testified to these instruments.

To prepare the MT reading comprehension test, the adapted reading comprehension test was translated by GT into Persian. The MT output was developed into a test of reading comprehension without making any

modifications. To check the practicality of the test, it was piloted. In the pilot study, 30 students took the test after being informed about the nature of the test and the purpose of their participation. Informing participants about the study they are taking part in is essential for ethical, autonomy, transparency, and data quality reasons. It helps to ensure that participants make informed decisions and that research is conducted in an ethical and transparent manner. The reliability index (KR-21) was calculated and the test was found to be desirably reliable ( $r = .78$ ). Similarly, to prepare the HT reading comprehension test, the very adapted reading comprehension test was translated into Persian by the researchers and then revised by one expert in translation. Few amendments were made to the test after pilot testing and collected scores were used to calculate the KR-21 index ( $r = .83$ ).

The researchers distributed the instruments (English Language Proficiency and Reading Comprehension Tests) and collected the quantitative data in two days. First, 211 university students were selected and took part in Quick Placement Test scores. Yet, 140 students who scored 2 SDs above and below the mean score were selected as the homogenized sample of the study who took the Reading comprehension tests later on.

The participants from each discipline were randomly divided into HT and MT groups of equal proficiency. The researchers clearly described the nature and content of the reading comprehension test they were taking and the purpose of its administration, as well. All the instructions were in Persian and the researchers patiently answered the questions posed by the participants before the administration. The conditions of the two administration sessions were strictly kept identical, especially in terms of time allotted to the students, examination setting, and time. The researchers read the test

instructions to the participants and clarified the ambiguities for the participants. They were also told how to answer the questions during the forty-minute test session. Furthermore, the participants were also informed that their performance would be kept confidential and would not be counted toward their final score. The participants in each group had both the original text and the translated Persian text as well so that they could consult the original text if they needed. However, it is worth mentioning that the researchers neither encouraged nor discouraged the participants to do so. That is, the availability of the original text was mainly for the sake of meeting authenticity criterion of testing rather than assisting the participants with the test tasks.

Posterior to test sessions the volunteers from the MT group took part in the one-to-one interview. The qualitative audio-recorded data from the interviews were transcribed and then analyzed using the content analysis method. Accordingly, the transcription was coded by the researchers and the emerged codes were re-examined after a two-week interval to increase the inter-coder reliability of the findings. Estimating inter-rater reliability involves defining the variable being rated, choosing an appropriate reliability coefficient, calculating the reliability coefficient, interpreting the result, and addressing sources of disagreement. By following these steps, it can be ensured that ratings are reliable and accurate. In addition, the researchers consulted two of their colleagues, experts in foreign language teaching, to review the transcription and the final coding draft of prepared by the researchers. Over 92 percent of the codes drawn from the transcription were approved by the reviewers and were incorporated for further classification of themes and interpretation of the findings.

## Results and Findings

### Addressing Research Question 1

Table 2 shows the results of the Spearman correlation test between their scores. It is worth mentioning that the Spearman correlation test was conducted because the distribution of the scores was not normal, as confirmed by the results of the Shapiro-Wilk test of normality ( $p < .05$ ).

The results reported in Table 2 show that the observed mean scores for the MT reading comprehension test ( $\bar{x} = 12.80$ ) and HT reading comprehension test scores ( $\bar{x} = 14.78$ ) were not very close. In addition, it has to be noted that the distributions of the HT and MT reading comprehension test scores of the students were not normal ( $p < .05$ ). Non-parametric tests were used to further analyze the results and answer the research questions. It can be argued that there was a significant direct moderate correlation ( $r = .59, p = .01 < .05$ ) between HT and MT reading comprehension test scores. Accordingly, it can be concluded that the participants' performance on one test can significantly moderately predict their performance on the other. The significance of the correlation between the overall HT and MT reading comprehension test scores implies that the MT output was comprehensible and,

therefore, enjoys an acceptable level of efficacy.

To further investigate the discrepancy between the HT and MT reading scores a pairwise comparison was made between the HT and MT reading comprehension test scores as observed for each type of comprehension skills described in the instrument section. The descriptive statistics are demonstrated below in Table 3, together with the results of the normality test and Mann-Whitney Test used for comparing the scores from the two groups. Because the distribution of the observed scores for both HT and MT reading comprehension sub-skill scores was not normal ( $p < .01$ ), the Mann-Whitney Test was used to compare the two groups.

As shown in Table 3 the statistics observed for the types of reading comprehension skills revealed that the distribution of the scores for all the sub-skills of reading comprehension, as measured by both HT and MT reading comprehension tests, is dispersed. Overall, the descriptive statistics imply that both HT output and MT output were similar with regard to their level of comprehensibility. Because the distribution of the observed scores for both HT and MT reading comprehension sub-skill scores was not normal ( $p < .01$ ), the Mann-Whitney Test was used to compare the two groups.

**Table 2: Results of and Spearman Correlation Test.**

		n	$\bar{x}$	SD	Shapiro-Wilk			Correlation	
					Statistic	df	p	r	p
Reading Scores	MT	70	12.80	3.31	.94	70	.00	.59	.01
	HT	70	14.78	2.27	.90	70	.02		

**Table 3: Descriptive Statistics for Each Type of Reading Comprehension Skills as Measured by the Reading Comprehension Test and the Normality of the Distribution**

	Group	Descriptive statistics			Shapiro-Wilk			Between-group comparison		Cohen's d
		n	$\bar{x}$	SD	Statistic	df	Sig.	U	p	
Scanning	MT	70	3.10	2.43	.904	70	.000	1.72	.08	.24
	HT	70	3.71	2.63	.833	70	.000			
Inference (Lexical)	MT	70	2.26	2.12	.781	70	.000	1.28	.23	.14
	HT	70	2.57	2.02	.798	70	.000			
Coherence	MT	70	1.06	1.00	.635	70	.000	.83	.31	.07
	HT	70	1.13	.95	.644	70	.000			
Skimming	MT	70	1.01	1.46	.672	70	.000	2.22	.02	.45
	HT	70	1.67	1.43	.685	70	.000			
Inference	MT	70	3.57	3.49	.643	70	.000	1.94	.19	.08
	HT	70	3.88	3.49	.632	70	.000			
Reference	MT	70	.51	.50	.636	70	.000	2.00	.01	.08
	HT	70	.55	.49	.632	70	.000			

The results shown in Table 3 demonstrated that there were significant differences between the HT and MT reading comprehension sub-skill scores in terms of skimming ( $U = 2.22, p = .02 < .05, d = .45$ ) and reference ( $U = 2.00, p = .01 < .05, d = .08$ ). However, those of the other measured sub-skills, scanning ( $U = 1.72, p = .08 > .05, d = .24$ ), inference (lexical) ( $Z = 1.28, p = .23 > .05, d = .14$ ), coherence ( $U = .83, p = .31 > .05, d = .07$ ) and inference ( $U = 1.94, p = .19 > .05, d = .08$ ) were insignificant. Moreover, considering the observed effect sizes, it can be concluded that the observed differences between all the measured sub-skills are negligible except the one for skimming which is moderate. Accordingly, it can be argued that the comprehensibility of MT output significantly lags behind that of HT in terms of skimming whereas HT output and MT output comprehensibility are significantly comparable in terms of scanning, inferencing a lexical item's meaning, coherence, reference, and inferencing the details.

### Addressing Research Question 2

To answer the second research question of the study, investigating how end MT users perceive the efficacy of the MT output for EAP reading comprehension, nine interviews were conducted with the participants and the collected data were further analyzed using the content analysis approach. The findings are elaborated below.

*Nahid: I occasionally use GT while reading an English passage, I mean, academic or non-academic. However, this was the first time I had a translated version in such a test and, if you ask me, it was helpful.*

*Mansoor: I believe that having a translated version of a passage is helpful. However, I think that relying solely on the translation could make it difficult for me to match the translated sentences with the original English text on the test, especially if I am unfamiliar with the topic.*

When examining the satisfaction level of students with the MT output in terms of its

comprehensibility, it should be noted that all nine interviewees in this study expressed satisfaction with the overall quality of the MT output. However, seven of the interviewees emphasized that the quality of the output was only acceptable when it was to be used for purposes other than the test they had taken. This is consistent with the second excerpt provided by Mansoor.

*Sadaf: It is not just about the translation we had on this test. Generally, the problem I have with such a translation is the grammatical errors, you know what I mean. For example, the [grammatical] suffixes of verbs in Persian may not be correct or you have two verbs in a sentence. You know, sometimes a verb is repeated twice. It is confusing sometimes.*

*Ali: Besides grammatical errors which are sometimes confusing, some technical words in the passage are not well-translated. As a matter of fact, I can understand it since I am familiar with the technical terms and I can fix it, for example, on this test, we had the original text and I could find the original term.*

In terms of MT output ambiguity, it has to be noted that as reported by the users, morpho-syntactic errors, as mentioned by six interviewees, and inappropriate lexical choices of the MT system, as highlighted by five interviewees, were the sources of ambiguity for the interviewees.

*Mina: The translation on this exam was the same as the translations I always use. To me, GT is a substitute for [bilingual] dictionaries. You know, it is better because I do not have to look up every word. It is faster since it translates paragraphs and I trust it because it never gives me several meanings for one word. I understand the translation.*

*It is enough for me that I can understand the text to answer the questions. It was enough for me in this test and it is enough for me in an English class or when I am reading a paper my teacher assigns.*

*Aref: Of course it is helpful. The point is that the translation is good enough for answering the critical questions a teacher may ask in an English class or .... Imagine I am going to be prepared for the [final] test. I have to ask my friends for further help or ask my teacher for further help. I even may have to ask for a professional [human] translation. However, GT has been a good substitute for all of them so far.*

In terms of the extent of MT usability for comprehending the texts, based on the findings from the interview, it has to be noted that eight interviewees believed that the output provides a reliable source of reference for the meaning of the general words and expressions and quick reference for inferring the general meaning, author's purpose, and the key points included in the test. However, it may not be enough to infer the implied meaning and critical understanding of the text.

## Discussion

---

The aim of the present study was to evaluate MT output based on the concept of translation efficacy defined based on its comprehensibility and usefulness for its target users. In addition, this study further aimed at probing the participants' perceptions of the efficacy of the MT output. The results of the study showed that the MT output is partially comparable to HT output in terms of its efficacy, which is defined as the degree of output comprehensibility and usefulness for the target users (readers). In addition, further investigation of the levels of comprehensibility of HT and MT output showed

that although overall comprehensibility of the MT and HT output is comparable, MT output is significantly inferior to HT output in terms of reference and skimming.

From the methodological perspective, this study, similar to previous ones such as Siregar [29], also addressed the use of reading comprehension tests for MT evaluation and proved comprehensibility to be a sound comprehensive criterion to be invested in for evaluating MT output. However, this research went beyond the mere reading comprehension test scores and encompassed target readers' perceptions about the efficacy of the MT output. This study revealed that cognitive criteria rather than linguistic ones such as readability, as emphasized by Bentivogli et al. [45] Cetiner and Isisag [35] and Doherty [46] of MT output has to play a central role in MT output evaluation so that translation quality in ESP courses had better to be replaced by translation efficacy which a relative measure of translation quality. That is, as indicated by the interviewees in this study, despite the fact that the MT output may lag behind a quality HT output, it may be sufficiently comprehensible and usable for target users. From the cognitive perspective of translation evaluation, it can be argued that the contribution of the target reader's familiarity and knowledge [47, 48, 49] and differences between the level of their comprehension and output may be best justified with regard to the contribution of reading comprehension strategies, both cognitive and metacognitive, to comprehension of texts in general and MT output, in particular [50].

The present study found that although the level of MT output comprehension of the target readers may be similar to that of the HT of the same source text, the comprehensibility of a given MT output is undeniably user-dependent. As implied in the previous studies [47, 48, 49],

this is to a great extent due to the fact that different target readers are employing different reading comprehension strategies, either cognitive or metacognitive, and are different in terms of their topic familiarity and background knowledge. That is, an MT with a specific translation quality may be of use for a group of users but may lack efficacy for a different group of users. However, considering some recent findings [51, 52, 53, 54, 55] it is unquestionable that the features of a text, like an MT output, influence the level of reading comprehension and, thus, comprehensibility of the text. According to aforementioned justifications, it can be argued that efficacy serves as a safe ground for evaluating the quality of translation; however, it has to be noted that it is both user-dependent context dependent in terms of the users' purposes which facilitates or limits the scope of readers' comprehension which in turn leads to rise or fall of MT output efficacy.

Concerning the results of the study reported above, there was an interesting finding considering the pairwise comparisons to which there was no significant difference between the participants' observed mean scores in the two groups in terms of inference, coherence, and scanning whereas there were significant differences in terms of skimming and reference. This may be best attributed to the inherent similarity of underlying cognitive comprehension processes occurring while reading MT and HT outputs, as described by Castilho and Guerberof Arenas [56]. Accordingly, both groups were significantly similar in terms of top-down components such as inference and coherence since (1) considering the interactive nature of reading comprehension, the participants in both groups had to stimulate similar schemata and rely on a similar body of background knowledge to run the required top-down cognitive processes; (2)

similar strategies were used to decode the MT output as those applied to comprehend the human translation output [50]. However, the lower mean score of the participants taking the MT test can be justified with regard to the difficulties imposed by erroneous sentences, idiomatic expressions, and stylistic distractions, which are inherent to machine-translated texts [41, 57]. Accordingly, it can be concluded that, based on the results of the study, while HT output has efficacy and serves the readers for a variety of academic purposes, MT output is limited in terms of its efficacy and serving the readers' academic purposes. In other words, it can be concluded that although MT output may not meet higher education students' academic needs comprehensively in terms of its efficacy with regard to skimming and reference, it certainly has efficacy similar to HT output in terms of satisfying their needs in terms of scanning, inference, and coherence.

## Conclusions

---

Translation and L1 use have been a commonplace teaching strategy in ESP courses held for university students in EFL contexts around the world. With the emergence of MT, students welcomed its use in and out of classes when encountering English texts. Although there also have been many studies supporting its use as an educational aid, its contribution to reading comprehension in ESP courses has been controversial due to its debatable translation quality. A review of the previous studies on evaluating translation quality indicates that this concept has been defined mostly as a text-dependent rather than a user-dependent concept so text features and textual errors of MT output have been widely researched. This study, however, adopted a different, relative, and pragmatic perspective to MT output

evaluation. Having inspired by previous studies, the author introduced the concept of translation efficacy which is mainly based on the comprehensibility and usefulness of the output for target users.

Comprehensibility, as an MT quality assessment yardstick, has been researched several times so far; however, its inherent relativity fits the concept of translation efficacy. That is, an MT output has efficacy as far as it is 'desirable' or 'good' enough for users to meet their specific reading purposes. The present study confirmed the fact that the efficacy of MT output is target-reader-dependent and text-dependent since it is determined both by the characteristics of the readers, such as their disciplines, and text features, as demonstrated by the significant differences in comprehension levels of the same readers measured by the same questions for HT and MT output. Accordingly, this study shed limelight on comprehensibility as a criterion of MT output efficacy; that is to say, it has to be reminded that MT quality needs to be defined as a context-bound and target-reader-specific concept.

The findings of this study feature some implications for MT users in academic contexts and ESP courses. Considering the user-dependency of MT, its end-users have to be born in mind and their characteristics such as strategic competence and background knowledge have been taken into account. In addition, it has to be remembered that with regard to text-dependency of the quality of MT output, its users have to notice the text type, content, genre before counting on an MT system since a given MT system may be suitable for translating a specific genre or content from a specific subject area, but not a different genre or content from another field of study. In addition, there is a need to consider MT literacy [58] by students and teachers. If MT is to be

widely used in English teaching courses, both teachers and students will ideally need some degree of MT literacy [59].

It is undeniable that the findings of the study have to be interpreted with caution since the number of participants in this study was low, and despite the reliability of the test, the number of reading comprehension sub-skills, items, and texts was not many. The next steps are to replicate this study with other languages than English as well as invite more participants. Further consideration may include texts of different genres and content.

### Authors' Contribution

Both authors were equally involved in conducting the study, analyzing the data, and writing the paper.

### Acknowledgments

The authors would like to thank all the students of Farhangian University who took part in the study.

### Conflicts of Interest

The authors have no conflict of interest.

### References

- [1] Tudor, I. Using translation in ESP. *ELT Journal*, 1987;41(4): 268-273. <https://doi.org/10.1093/elt/41.4.268>
- [2] Plastina, A. F. Mental translation in the ESP classroom: a case study on processing reading comprehension. *Mental Translation in the ESP Classroom*, 2006; 1000-1013.
- [3] Groves, M., & Mundt, K. Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, 2015; 37: 112-121.
- [4] Bahri, H., & Mahadi, T. Google Translate as a supplementary tool for learning Malay: A case study at Universiti Sains Malaysia. *Advances in Language and Literary Studies*, 2016; 7(3): 161–167. <http://dx.doi.org/10.5755/j01.sal.0.21.1469>
- [5] Lee, S. M. The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, 2020; 33(3): 157-175. <https://doi.org/10.1080/09588221.2018.1553186>
- [6] Tsai, S. C. Using google translate in EFL drafts: a preliminary investigation. *Computer Assisted Language Learning*, 2019; 32(5-6): 510-526. DOI:10.1080/09588221.2018.1527361
- [7] Suhono, S., Zuniati, M., Pratiwi, W., & Hasyim, U. A. A. Clarifying Google Translate Problems Of Indonesia-English Translation Of Abstract Scientific Writing. *Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA)*. 2020; Padang, Indonesia. <http://dx.doi.org/10.4108/eai.24-1-2018.2292399>
- [8] Knowles, C. L. Investigating instructor perceptions of online machine translation and second language acquisition within most commonly taught language courses. [dissertation]. The University of Memphis; 2016.
- [9] Rushwan, I. M. H. The role of translation in developing ESP learners' reading comprehension skills-A case study of medical students at Najran University-ksa. *International Journal of Applied Linguistics and English Literature*, 2017; 6(3): 243-253. <https://doi:10.7575/aiac.ijalel.v.6n.3p.243>
- [10] Mesbah, M., & Faghani, M. Task-based and grammar translation teaching methods in teaching reading comprehension to nursing students: An action research. *Aula Orientalis*, 2015; 1: 319-25.
- [11] Avand, A. Q. Using translation and reading comprehension of ESP learners. *The Asian ESP Journal*, 2009; 5(1): 44-60.
- [12] Kavaliauskien?, G. Proficiency in reading, writing and translation skills: ESP aspect. *Vertimo Studijos*, 2009; 2(2): 171-184.
- [13] Malekan, F., & Hajimohammadi, R. The Relationship between Iranian ESP Learners' Translation Ability and Resilience in Reading Comprehension. *International Journal of Education and Literacy Studies*, 2017; 5(2): 47-52.
- [14] Pham, C. T. K. Reading comprehension and translation performance of English linguistics students of Hung Vuong University: A correlational study. *International Journal of English Language and Translation Studies*, 2017; 5: 79-85.
- [15] Tavakoli, M., Shafiei, S., & Hatam, A. H. The relationship between translation tests and reading comprehension: a case of Iranian University students. *Iranian Journal of Applied Language Studies*, 2012; 4(1): 193-211. <https://doi:10.22111/IJALS.2012.1353>
- [16] Laviosa, S., & Cleverton, V. Learning by translating: A contrastive methodology for ESP learning and translation. *Scripta Manent*, 2015; 2(1): 3-12.
- [17] Leonardi, V. Teaching business English through translation. *Journal of Universal Language*, 2009; 10(1): 139-153. <https://doi:10.22425/jul.2009.10.1.139>
- [18] Mažeikiene. VTranslation as a method in teaching ESP: An inductive thematic analysis of literature. *Journal of Teaching English for Specific and Academic Purposes* .2019; 6(3): 513-523. <https://doi.org/10.22190/JTESAP1803513M>



- [19] Novita, L., & Mustafha, B. Translation as a Pedagogical Device for Improving Students' Reading Comprehension in the ESP text. In Proceedings of the 1st International Conference on Business, Law And Pedagogy, ICBLP 2019. *European Alliance for Innovation (EAI)*.  
<http://dx.doi.org/10.4108/eai.13-2-2019.2286060>
- [20] Levy, M., & Steel, C. Language learner perspectives on the functionality and use of electronic language dictionaries. *ReCALL: the Journal of EUROCALL*, 2015; 27(2): 177. <https://doi.org/10.1017/S095834401400038X>
- [21] Lee, S. M., & Briggs, N. Effects of using machine translation to mediate the revision process of Korean university students' academic writing. *ReCALL*, 2021; 33(1): 18-33.  
<https://doi.org/10.1017/S0958344020000191>
- [22] Nino, A. Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, 2008; 21(1): 29-49.  
<http://doi.org/10.1080/09588220701865482>
- [23] Melby, A. K. The translation profession and the computer. *CALICO Journal*, 2013; 1(1): 55-57.  
<https://doi.org/10.1558/cj.v1i1.55-57>
- [24] Garcia, I., & Pena, M. I. Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, 2011; 24(5): 471-487.  
<https://doi.org/10.1080/09588221.2011.582687>
- [25] Chung, E. S., & Ahn, S. The effect of using machine translation on linguistic features in L2 writing across proficiency levels and text genres. *Computer Assisted Language Learning*, 2021; 5(3): 1-26. <https://doi.org/10.1080/09588221.2020.1871029>
- [26] Chang, C. K., & Hsu, C. K. A mobile-assisted synchronously collaborative translation-annotation system for English as a foreign language (EFL) reading comprehension. *Computer Assisted Language Learning*, 2011; 24(2): 155-180.  
<https://doi.org/10.1080/09588221.2010.536952>
- [27] Arnold, D., & Balkan, L. *Machine Translation*. 2001; USA: Blackwell Publishers.
- [28] Alhaisoni, An investigation of Saudi EFL university students' attitudes towards the use of google translate. *International Journal of English Language Education*, 2017; 5(1): 72-82.  
<https://doi.org/10.5296/ijele.v5i1.10696>
- [29] Siregar, R. Exploring the undergraduate student's perception on translation - A preliminary step to teach translation in EFL classes. *English Language Teaching*, 2018; 11(9): 90-101. DOI:10.5539/elt.v11n9p90
- [30] Shterionov, D., Superbo, R., Nagle, P., Casanellas, L., O'dowd, T., & Way, A. Human versus automatic quality evaluation of NMT and PBSMT. *Machine Translation*, 2018; 32(3): 217-235.
- [31] Taleghani, M., Pazouki, E., & Ghahraman, V. The Correlation of Machine Translation Evaluation Metrics with Human Judgement on Persian Language. *Journal of Language and Translation*, 2019; 9(3): 43-55.  
<https://doi.org/10.1001.1.20088590.2019.9.3.4.6>
- [32] Bahdanau, D., Cho, K., & Bengio, Y. Neural machine translation by jointly learning to align and translate. 2014; Retrieved online on February 23, 2020 from arXiv preprint arXiv:1409.0473.
- [33] Napitupulu, S. Analyzing Indonesian-English abstracts translation in view of translation errors by google translate. *International Journal of English Language and Linguistics Research*, 2017; 5(2): 15-23.
- [34] Salah Habeeba, & Muhammed Investigating the effectiveness of google translate among Iraqi students. *International Journal of Innovation, Creativity and Change*, 2019; 7(8): 112-122.
- [35] Cetiner, C., & Isisag, K. Undergraduate level translation students' attitudes towards machine translation post-editing training. *International Journal of Languages' Education and Teaching*, 2019; 7(1): 110-120. <https://doi.org/10.18298/ijlet.3242>
- [36] Koehn, P., & Knowles, R. Six challenges for neural machine translation. 2017; Retrieved online on February 23, 2020 from arXiv preprint arXiv:1706.03872.
- [37] Chang, C., & Yamada, M. Translation tasks for learning collocations: Effects of machine translation plus post-editing and sight translation. *English Teaching and Learning*, 2020; 3(5): 87-101. <http://doi.org/10.1007/s42321-020-00059-x>
- [38] Toral, A., & Sánchez-Cartagena, V. M. A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions. 2017; Retrieved online on February 23, 2020 from arXiv preprint arXiv:1701.02901.
- [39] Jimenez, C. The role of translation technologies in Spanish language learning. *Journal of Spanish Language Teaching*, 2018; 4(4): 141-160.  
<http://doi.org/10.1080/23247797.2017.1408949>
- [40] Abdelaal, N., & Alazzawie, A. Machine Translation: The case of Arabic-English translation of news texts. Theory and Practice in Language Studies, 2020; 10(4): 408-418.  
<http://doi.org/10.17507/tpls.1004.09>
- [41] Roturier, J. An investigation into the impact of controlled English rules on the comprehensibility, usefulness and acceptability of machine-translated technical documentation for French and German users. [dissertation]. Dublin City University; 2006.
- [42] Alderson, C. *Reading Assessment*. 2000; New York: Cambridge University Press.
- [43] Lewkowicz, J. A. Authenticity in language testing: some outstanding questions. *Language testing*, 2000; 17(1): 43-64.
- [44] Bachman, L. F., & Palmer, A. S. *Language Testing in Practice: Designing and Developing Useful Language Tests*. 1996; Oxford: Oxford University Press.

- [45] Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. Neural versus phrase-based machine translation quality: a case study. 2016; Retrieved online on February 23, 2020 from arXiv preprint arXiv:1608.04631.
- [46] Doherty, S. Investigating the effects of controlled language on the reading and comprehension of machine translated texts: A mixed-methods approach [dissertation]. Dublin City University; 2012.
- [47] Anuyahong, B. A study of English reading strategies and reading motivation used by Thai-Nichi institute of technology students. *Automotive Engineering*, 2011; 1(14): 21-30.
- [48] Becirovic, S., Brdarevic-Celjo, A., & Sinanovic, J. The Use of Metacognitive Reading Strategies among Students at International Burch University: A Case Study. *European Journal of Contemporary Education*, 2017; 6(4): 645-655. <https://doi.org/10.13187/ejced.2017.4.645>
- [49] Ganji, M., Yarahmadzahi, N., & Sasani, N. English Major Students' Awareness of Metacognitive Reading Strategies: Gender and Academic Level in Focus. *Issues in Language Teaching*, 2018; 7(2): 91-119. <https://doi.org/10.22054/ilt.2019.46700.428>
- [50] Karnal, A. R., & Pereira, V. V. Reading strategies in a L2: A study on machine translation. *Reading Matrix: An International Online Journal*, 2015; 15(2): 69-80.
- [51] Brantmeier, C. Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish. *The Modern Language Journal*, 2005; 89(1), 37-53. <https://doi.org/10.1111/j.0026-7902.2005.00264.x>
- [52] Chen, C. M., & Lin, Y. J. Effects of different text display types on reading comprehension, sustained attention and cognitive load in mobile reading contexts. *Interactive Learning Environments*, 2016; 24(3): 553-571. <https://doi.org/10.1080/10494820.2014.891526>
- [53] Coelho, C. L. G., & Correa, J. Reading comprehension: Cognitive abilities and types of text. *Psico*, 2017; 48(1): 40-49.
- [54] Matheson, I. Unpacking Reading Comprehension by Text Type: An Examination of Reading Strategy Use and Cognitive Functioning in Poor and Typically-Achieving Comprehenders [dissertation]. Queen's University, Kingston; 2018.
- [55] Uysal, P. K., & Bilge, H. An investigation on the relationship between reading fluency and level of reading comprehension according to the type of texts. *International Electronic Journal of Elementary Education*, 2019; 11(2): 161-172. <https://doi.org/10.26822/iejee.2019248590>
- [56] Castilho, S., & Guerberof Arenas, A. Reading comprehension of machine translation output: what makes for a better read? In P´erez-Ortiz, S´anchez-Mart´nez, Esp´a-Gomis, Popovi´c, Rico, Martins, Van den Bogaert, Forcada (eds.). 2018; Proceedings of the 21st Annual Conference of the European Association for Machine Translation 79–88.
- [57] Scarton, C., & Specia, L. A reading comprehension corpus for machine translation evaluation. 2016; In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 3652-3658.
- [58] O'Brien, S., & Ehrensberger-Dow, M. MT Literacy—A cognitive view. *Translation, Cognition & Behavior*, 2020; 3(2): 145-164. <https://doi.org/10.1075/tcb.00038.obr>
- [59] Bowker, L. Machine translation literacy instruction for international business students and business English instructors. *Journal of Business & Finance Librarianship*, 2020; 25(1-2): 25-43. DOI:10.1080/08963568.2020.1794739
- [60] Chen, S. (2019). Effects of machine translation on foreign language learning and teaching. *Journal of Language Teaching and Research*, 10(5): 991-1000. <https://doi.org/10.17507/jltr.1005.13>
- [61] Li, D., & Liu, J. (2020). The impact of machine translation on foreign language learning. *Open Journal of Linguistics*, 10(2): 255-262. <https://doi.org/10.4236/ojml.2020.102016>
- [62] Liu, J., & Li, D. (2020). The impact of machine translation on foreign language teaching. *International Journal of Emerging Technologies in Learning*, 15(11): 143-159. <https://doi.org/10.3991/ijet.v15i11.14627>

## AUTHOR(S) BIOSKETCHES

**Vahid R. Mirzaeian** has got a PhD. in CALL from the University of Manchester. He is currently working on CALL-related issues.

**Mirzaeian, V. Associate Professor, Applied and Computational Linguistics, Alzahra University, Tehran, Iran**

✉ [mirzaeian@alzahra.ac.ir](mailto:mirzaeian@alzahra.ac.ir)

**Mojtaba Maghsudy** got his PhD. from Puna University in India. He is currently an academic member of staff at Farhangian University working on areas related to foreign language learning.

**Maghsoudi, M. Associate Professor, Department of English Language Teaching, Farhangian University, Tehran, Iran**

✉ [Maghsduim@cfu.ac.ir](mailto:Maghsduim@cfu.ac.ir)

**Citation (Vancouver):** Mirzaeian V, Maghsoudi, M. [Benchmarking Machine Translation Efficacy for Teaching EAP Reading Comprehension Skills]. *Tech. Edu. J.* 2024; 18(1): 1-18

 <https://doi.org/10.22061/tej.2023.9892.2913>

